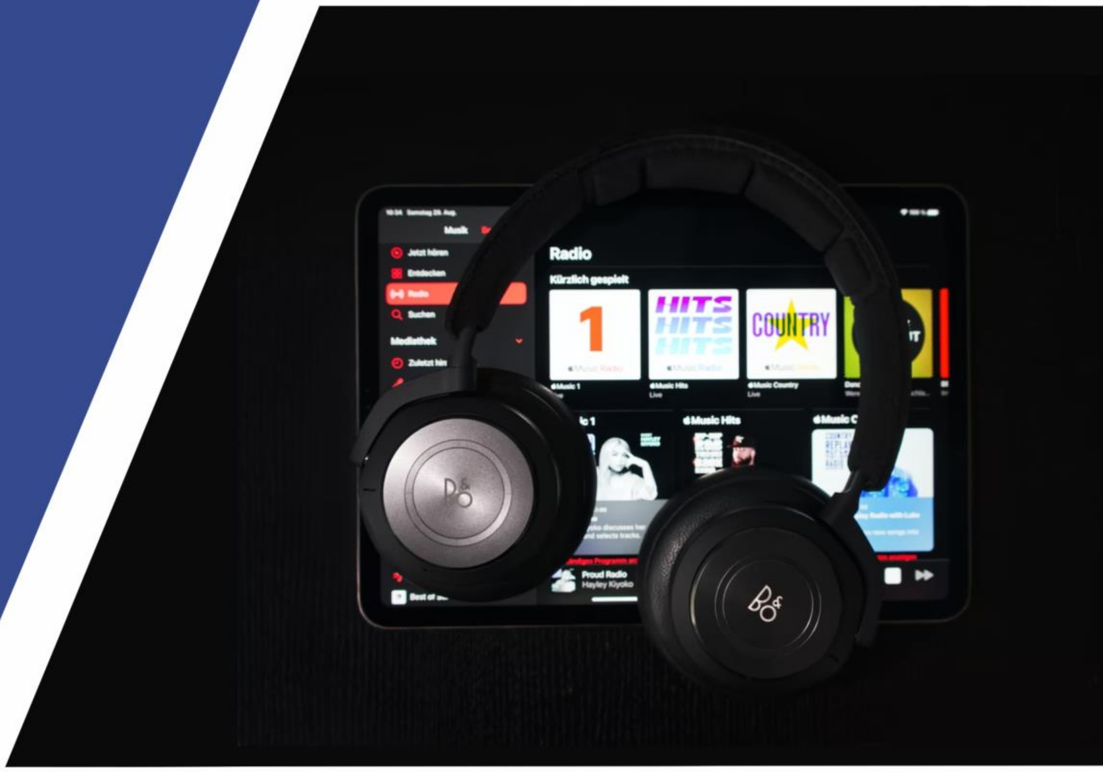
NATIONAL OPEN UNIVERSITY OF NIGERIA

CIT 891

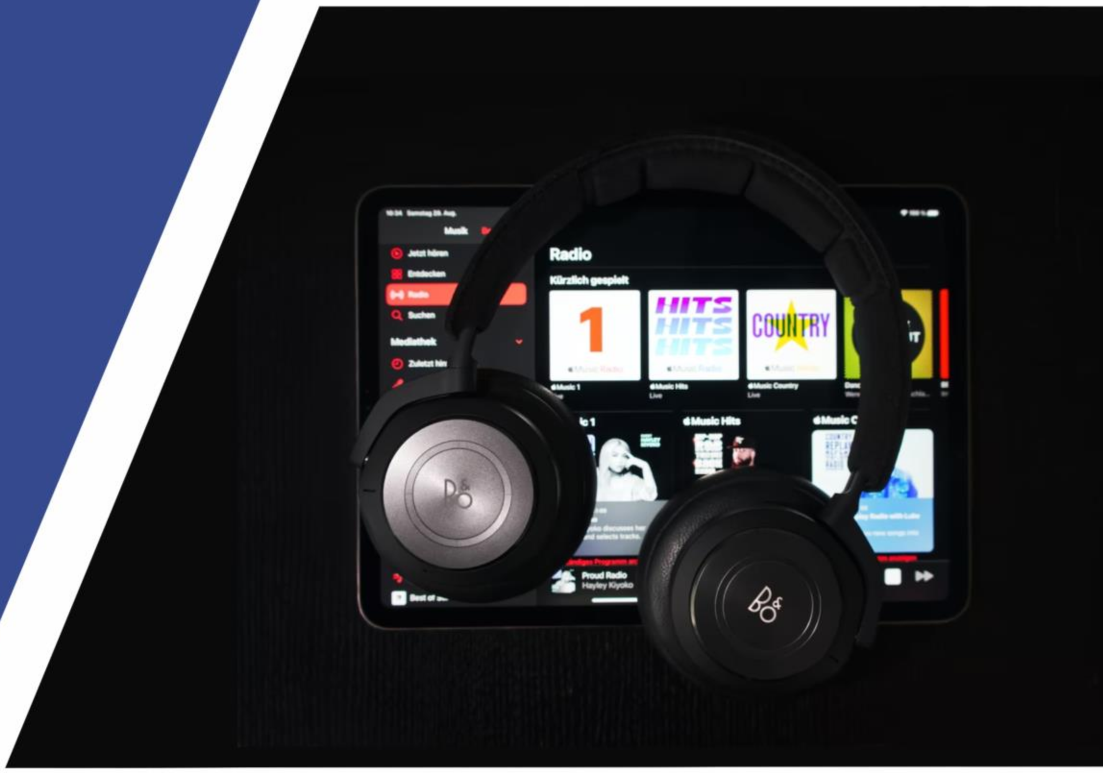


Advanced Multimedia Technology

Module 4

NATIONAL OPEN UNIVERSITY OF NIGERIA

CIT 891



Advanced Multimedia Technology

Module 4

CIT 891
Advanced Multimedia Technology
Module 1

Course Developer/Writer
Dr. Ikhu-Omoregbe, Nicholas, National Open University of Nigeria

Course Coordinator
A. A. Afolorunso, National Open University of Nigeria

Programme Leader
Prof. MoniOluwa Olaniyi, National Open University of Nigeria

Credits of cover-photo: Henry Ude, National Open University of Nigeria

National Open University of Nigeria - University Village, Plot 91, Cadastral Zone, Nnamdi Azikiwe Expressway, Jabi, Abuja-Nigeria

www.nou.edu.ng centralinfo@nou.edu.ng
oer.nou.edu.ng oerunit@noun.edu.ng OER repository
Published in 2013, 2021 by the National Open University of Nigeria

How to re-use and attribute this content

Unit 1 Image Coding With Dct and Wavelet Technologies

1.0 Introduction

The discrete cosine transform (DCT) is widely applied in image compression. Alternative transforms such as Walsh, Hadamard, Karhunene-Loeve and wavelets have also been introduced. With wavelets the basic functions are not restricted to sine waves and can be quite different in form and thus characteristics. By using wavelets we get better image quality at low data rates and higher compression when compared to the DCT-based compression.

2.0 Objectives

At the end of this unit, you should be able to:

- Explain image coding with DCT
- Discuss the application of wavelet technology in image coding.

3.0 Main Content

3.1 Image Coding with DCT

One of the most common data transformations used in video compression is the discrete cosine transformation (DCT). It is in CCITT H.261, JPEG, H.320 group of standards for video conferencing, and in both the MPEG1 and MPEG 2 standards. The discrete cosine transform is a special case of the Fourier transform applied to discrete (sample) signal, which decomposes a periodic signal into a series of sine and cosine harmonics functions. The signal can then be represented by a series of coefficient of each of these functions.

We provided some mathematical foundation in Module two. Under certain conditions, the DCT separates the signal into only one series of harmonic cosine function in phase with the signal, which reduces by half the number of coefficient necessary to describe the signal compared to a Fourier transform.

In the case of picture, the original signal is a sampled bidimensional DCT (horizontal and vertical directions), which will transform the luminance (or chrominance) discrete value of a block N X N pixels into another block (or matrix) of N X N coefficient representing the amplitude of each of the cosine harmonic functions.

The transformed block consist value of horizontal and vertical frequencies. In the block coefficients of the horizontal axis represent increasing horizontal frequencies from left to right, and on the vertical axis they represent increasing vertical frequencies from top to bottom. The very first coefficient in the top left corner that is coordinates (0,0) denotes null horizontal and vertical frequencies, and is therefore called the **DC** coefficient, and the bottom right coefficient represent the highest spatial frequency component in the two direction.

In order to carry out this exercise, a block size of 8 x 8 pixels (see figure 3.1) which the DCT transforms into a matrix of 8 x 8,(see figure 3.2) ,is generally used.

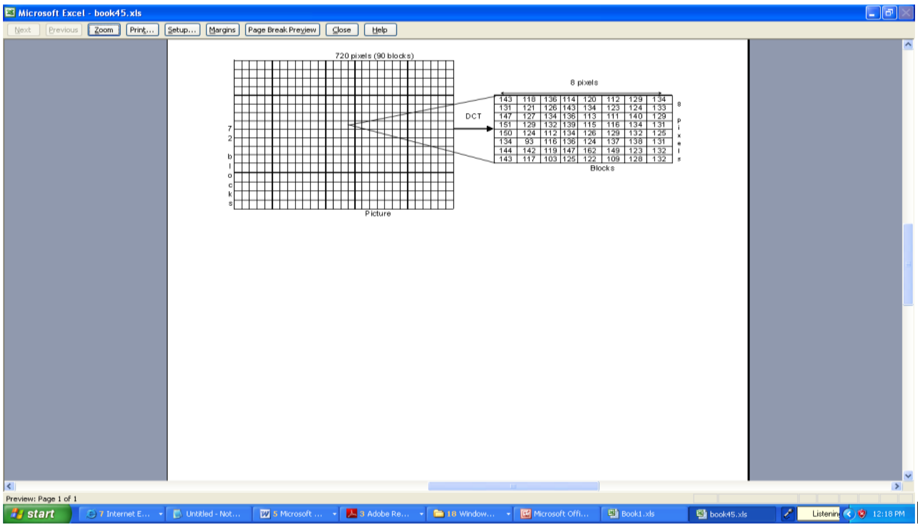


Fig.3.1 : Cutting of Blocks of 8 x 8 Pixels

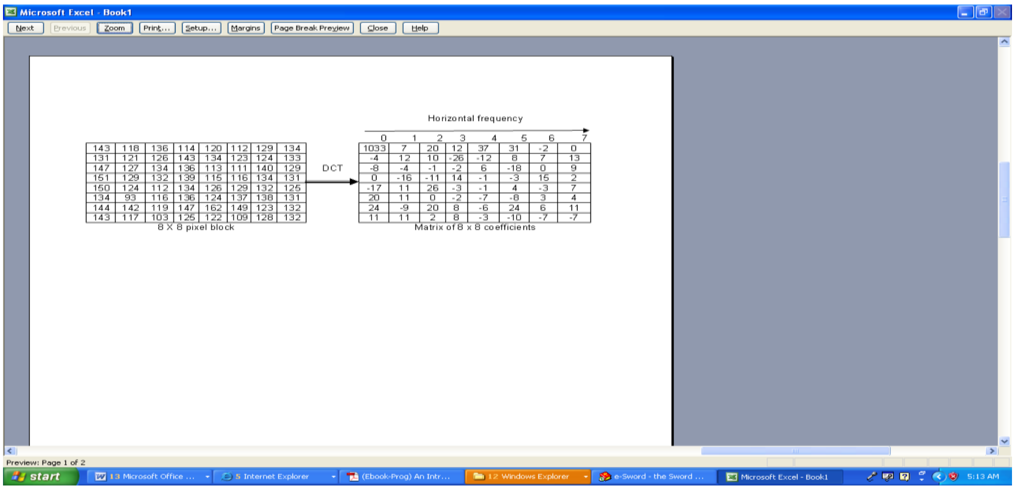


Fig 3.2: Transformation of a Block of 8 x 8 Pixels into a Matrix of 8 x 8 Coeff. Using DCT

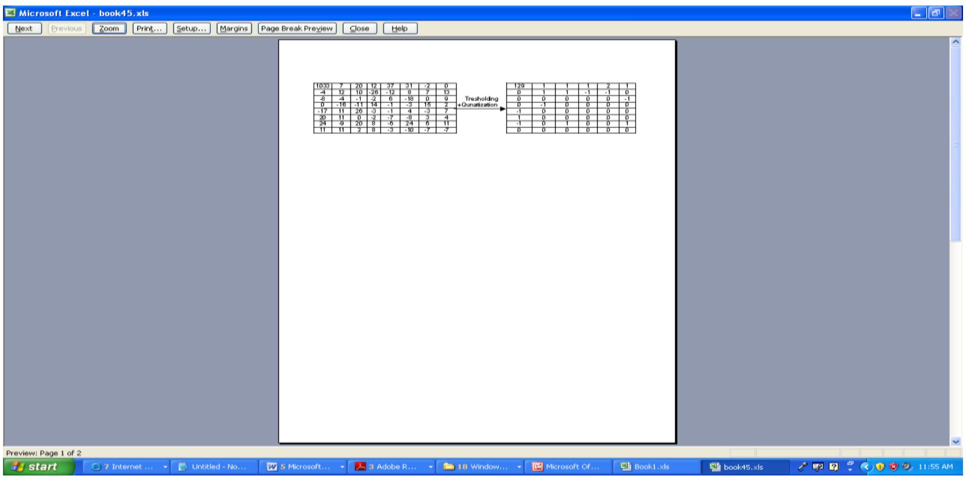


Fig. 3.3: Result of Thresholding and Quantisation

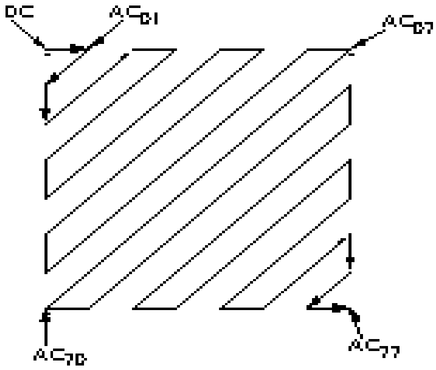


Fig. 3.4: Zigzag Reading of the Coefficient of the Matrix

The DCT has notable property of concentrating the energy of the block on a comparatively low number of coefficient situated in the top left corner of the matrix with coefficient that are decorated from each other. This property is taken advantage of in subsequent compression processes. The DCT transform process is reversible. Nevertheless, due to the nature of human vision (reduced sensitivity to high spatial frequencies), it is possible, without perceptible degradation of the picture quality, to eliminate the values below a certain threshold function of the frequency. The laminated values are replaced by 0 (an operation known as thresholding); this part of the process is obviously not reversible, as some data are thrown away.

The remaining coefficients are then quantised with an accuracy decreasing with the increasing spatial frequencies, this further reduces the quantity of information required to encode a block. This process is also not reversible, but it has little effect on the perceived picture quality. The thresholding /quantisation process is depicted in figure 3.3

In order to regulate the bit-rate required to transmit moving pictures, the thresholding and quantisation parameter are used to dynamically regulate the bit-rate for moving pictures. A serial bitstream is then obtained by "zig-zag" reading of the coefficients, as shown in Fig 3.4.to allow a relatively long series of null coefficient to be obtained as quickly as possible, in order to increase the efficiency of the compression algorithm (such as, Run length, Huffman) used.

3.2 Application of DCT

DCT is used for most image/ video standards. Two popular examples of such standards are JPEG (Joint photographic expert group) and MPEG (Motion picture expert group)

a. The joint photographic expert group (JPEG) is a standard for compression of still pictures. The colour signal, red, green and blue are sampled and each color component is transformed by DCT in 8 x 8 pixel blocks. The DCT coefficients are quantised and encoded in a way that the more important lower components are represented by more bits than the higher-frequency coefficients. The coefficients are reordered by reading the DCT coefficient matrix in a zigzag fashion, and the data stream is Huffman coded.

b. The motion picture expert group (MPEG) standard MPEG-1, for compression of full-motion pictures on digital storage media such as CD-ROM and digital versatile disc (DVD), with a bit transfer rate of about 1.5 Mbits/s is similar to JPEG. With this standard compression is achieved by allowing sampled framed split into blocks to be transformed using DCT in the same way as JPEG. The coefficients are then coded with either forward or backward prediction or a combination of both. The output from the predictive coding is then quantised using matrix of quantisation steps. Since MPEG is more complicated than JPEG, it requires even more computing power. We shall discuss the MPEG standard in more details in the unit 3 of this module.

3.3 Wavelets Compression

Wavelets first appeared in the 1980s in geophysical work but had long spread to other areas such as mathematics, computer science and engineering. The basic idea behind it application is that many signals can be represented by combining many simpler wave forms called basic functions by using weighing factors known as co-efficient. Similar to this is the Fourier transform which will decompose a signal into a set of sine waves of specific frequency. With wavelets the basic functions are not restricted to sine waves and can be quite different in form and thus characteristics.

Representation of a signal using sinusoids is very effective for stationary signals, which are statistically predictable and are time-invariant in nature. Wavelet representation is found to be very effective for nonstationary signals, which are not statistically predictable and time-varying in nature. Variation of intensity to form edges is a very important visual characteristic of an image. From signal theoretic perspective, discontinuities of intensities occur at the edges in any image and hence it can be prominently visualized by the human eye. The time and frequency localization property of wavelets makes it attractive for analysis of images because of discontinuities at the edges.

Wavelets are functions defined over a finite interval and having an average value of zero. The basic idea of the wavelet transform is to represent any arbitrary function (t) as a superposition of a set of such wavelets or basis functions. These basis functions also referred to as baby wavelets are obtained from a single prototype wavelet called the mother wavelet, by dilations or contractions (scaling) and translations (shifts).

If the mother wavelet is denoted by $\psi(t)$ the other wavelets $\psi^{a,b}(t)$ for a > 0 and a real number *b* can be represented as:

$$\psi^{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$$

where a and b represent the parameters for dilations and translations in the time domain. The parameter a causes contraction in time domain when a < 1 and expansion when a > 1. Figure 3.5 is used to illustrate a mother wavelet and its contraction and dilation.

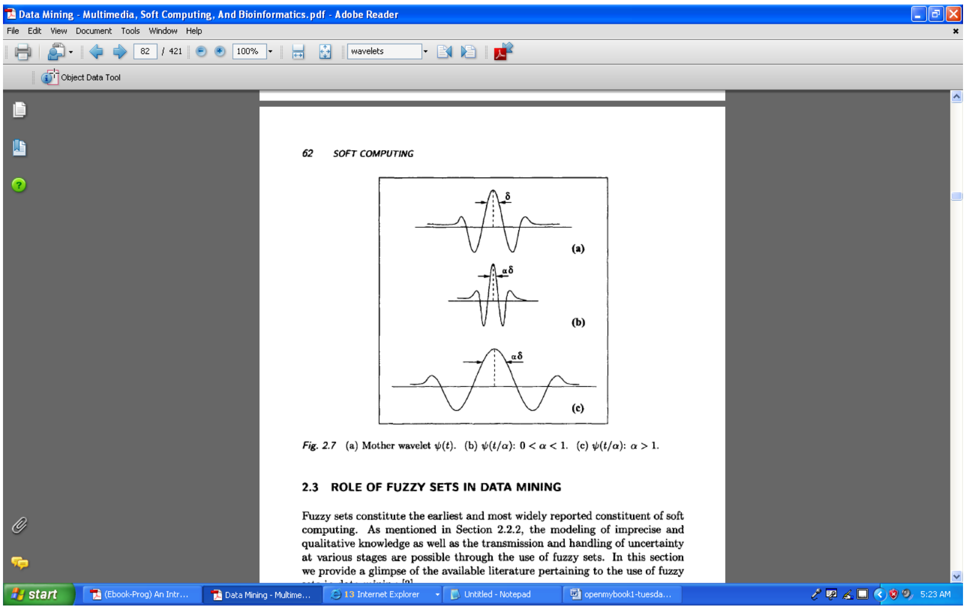


Fig. 3.5: (a) Mother Wavelet $\psi(t)$ **(b)** $\psi(t/\alpha):0<$ **α<1 (c)** $\psi(t/\alpha):\alpha>1$

By using wavelets we get better image quality at low data rates and higher compression when compared to the DCT-based compression.

Wavelets compression is in some aspects very similar to the DCT compression techniques that are used with MPEG and H.320 video conferencing. It transforms the data to make it easier to decide which data can be lost for the minimum impact on video quality and then uses coding techniques to compress the data. The big difference though is that the DCT algorithm is not used and is replaced with wavelet transform. Using wavelets gives better image or video quality at low data rates and higher compression when compared to the DCT-based compression.

Compression schemes based on DCT requires that the input image needs to be ``blocked", thus correlation across the block boundaries is not eliminated. This results in noticeable and annoying ``blocking artifacts" particularly at low bit rates as shown in figure 3.6(a). Blocking artifacts also called macroblocking are distortion that appears in compressed image / video material as abnormally large pixel blocks. It occurs when the encoder cannot keep up with the allocated bandwidth. It is especially visible with fast motion sequences or quick scene changes.

Video uses lossy compression, and the higher the compression rate, the more content is removed. At decompression, the output of certain decoded blocks makes surrounding pixels appear averaged together and look like larger blocks. As TVs get larger, blocking and other artifacts become more noticeable.

Since there is no need to block the input image and its basis functions have variable length, wavelet coding schemes at higher compression avoid blocking artifacts. Wavelet-based coding is more robust under transmission and decoding errors, and also facilitates progressive transmission of images. In addition, they are better matched to the HVS characteristics. This is because of their inherent multiresolution nature. Wavelet coding has earlier been discussed in module two. Wavelet coding schemes are especially suitable for applications where scalability and tolerable degradation are important such as in Subband Coding.

The fundamental concept behind Subband Coding (SBC) is to split up the frequency band of a signal and then to code each subband using a coder and bit rate accurately matched to the statistics of the band. SBC has been used extensively first in speech coding and later in image coding because of its inherent advantages namely variable bit assignment among the subbands as well as coding error confinement within the subbands.

Fig. 1.6(a)

Fig. 1.6(b)

Figure 1.6(a) Original Lena Image, and (b) Reconstructed Lena with DC Component only, to show Blocking Artifacts.

Self -Assessment Exercise

i. What are blocking artifacts?
ii. What are wavelets?

4.0 Conclusion

The basic idea of transform compression is to extract appropriate statistical properties, for instance Fourier coefficients, of an image and let the most significant of these properties represent the image. The image is then reconstructed (decomposed) using an inverse transform. Often it is convenient to express the transform coefficient as a matrix. Two popular transforms that can be used are discrete cosine transform and discrete wave transform.

While the DCT-based image coders perform very well at moderate bit rates, at higher compression ratios, image quality degrades because of the artifacts resulting from the block-based DCT scheme. Wavelet-based coding on the other hand provides substantial improvement in picture quality at low bit rates because of overlapping basis functions and better energy compaction property of wavelet transforms.

5.0 Summary

Image and video compression have become an integrated part of today's digital multimedia applications. In this unit, we covered some of the more sophisticated techniques that take advantage of the statistics of the wavelet coefficients.

Wavelet theory is also a form of mathematical transformation, similar to the Fourier transform in that it takes a signal in time domain, and represents it in frequency domain. Wavelet functions are distinguished from other transformations in that they not only dissect signals into their component frequencies, they also vary the scale at which the component frequencies are analysed. Therefore wavelets, as component pieces used to analyse a signal, are limited in space. The ability to vary the scale of the function as it addresses different frequencies also makes wavelets better suited to signals with spikes or discontinuities than traditional transformations such as the Fourier transform. Applications of wavelet theory include JPEG2000 which is based on wavelet compression algorithms.

6.0 Self-Assessment Exercise

1. What are the advantages of wavelets over DCT?
2. Briefly describe two areas of application of DCT
3. What are wavelets?

7.0 References/Further Reading

Benoit,H. (1996). *Digital Televison MPEG-1, MPEG-2 and Principles of the DVB systems*.Oxford: Focal Press, Linacre House, Jordan Hill.

Gibson,J.D. (Ed.). (2001). *Multimedia Communications: Directions and Innovations* USA: Academic Press.

Strannedby,D., Walker,W.(2004). *Digital Signal Processing and Applications Newnes*.Oxford: An imprint of Elsevier, Jordan Hill.

Unit 2 Video Coding With Motion Estimation

1.0 Introduction

Digital video has become mainstream and is being used in a wide range of applications including DVD, digital TV, HDTV, video telephony, and teleconferencing. These digital video applications are feasible because of the advances in computing and communication technologies as well as efficient video compression algorithms. Most of the video compression standards are based on a set of principles that reduce the redundancy in digital video. In this unit we explained motion estimation by discussing video coding techniques with MPEG-1.

2.0 Objectives

At the end of this unit, you should be able to:

- explain the meaning of motion estimation and compensation
- explain the different types of frame
- describe the principles behind MPEG-1 Video Coding.

3.0 Main Content

3.1 What is a Digital Video?

We introduced the principles of video compression in module two. Digital video is essentially a sequence of pictures displayed overtime. Each picture of a digital video sequence is a 2D projection of the 3D world. Digital video thus is captured as a series of digital pictures or sampled in space and time from an analog video signal. A frame of digital video or a picture can be seen as a 2D array of pixels. We shall proceed to discuss video compression with respect to MPEG.

3.2 Video Coding of MPEG-1

The main objective of MPEG-1 was to reach a medium quality video with a constant total bit-rate of 1.5 Mbps for storing video and audio on CD-ROM. The video part used 1.15 Mbps, while the remaining 350 kb/s is used by audio and other details required by the system. Nevertheless the MPEG-1 specification is such that it allows different parameters to be chosen depending on the compromise between encoder complexity, compression rate and quality.

The MPEG-1 standard allows the user to set a whole range of parameter which controls the image size, target bitstream rate, etc. However, in order to assure interoperability, a constrained parameter set has been defined and most MPEG-1 decoders conform to this. The implication of this is that, for parameters outside this range, there is less likelihood of compatibility between different encoders and decoders.

Horizontal resolution less or equal 760 pixels

Vertical resolution less or equal 576 pixels

- Macroblock per picture less or equal 396
- Macroblocks to be processed per second less or equal 99000
- Frames per second less or equal 30
- Bitstream bandwith less or equal 1.86 Mbps
- Decoder buffer size less or equal 376832 pixels

These techniques, referred to as "prediction with movement compensation", consist of deducing most of the pictures of a sequence from preceding and even subsequent pictures, with minimum of additional information representing the differences between pictures. This usually requires an additional task of movement estimator in the MPEG encoder. This happens to be the greatest task and goes a long way in determining the encoder's performance. Fortunately, this function is not required in the decoder.

When dealing with video, we are actually talking about moving pictures in which case, decoding has to be accomplished in real time i.e. it has to be done in an acceptable and constant processing delay. This can only be accomplished, for the time being at least through, some specialised hardware. The coding, is usually a more complex process and can be done in more than one "pass" for applications where real time is not required but where quality is paramount (such as in engraving of disks); real time will, however, be required for many applications, such as 'live' video transmissions.

In the practical realisation of the encoder, a trade-off among speed, compression rate, complexity and picture quality is necessary. In addition, synchrononisation time and random access time to a sequence have to be maintained within an acceptable limit (not exceeding 0.5 s), which restricts the maximum number of pictures that can be dependent on the first picture to between 10 and 12 for a system operating at 25 fps(frames per second)

3.2.1. The different types of MPEG Frames

MPEG defines three types of pictures which are arranged as shown in Figure 3.2.1a

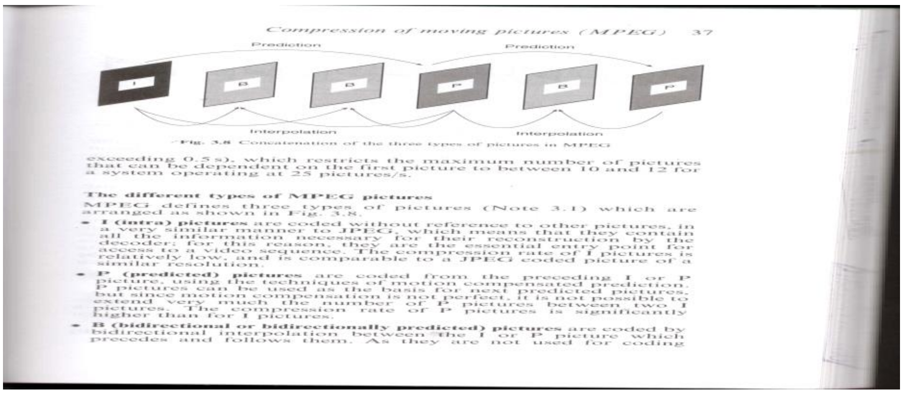

Compression of moving pictures (MPEG) 37

Fig. 3.8 Concatenation of the three types of pictures in MPEG

exceeding 0.5 s), which restricts the maximum number of pictures that can be dependent on the first picture to between 10 and 12 for a system operating at 25 pictures/s.

The different types of MPEG pictures

MPEG defines three types of pictures (Note 3.1) which are arranged as shown in Fig. 3.8.

- **I (intra) pictures** are coded without reference to other pictures, in a very similar manner to JPEG, which means that they contain all the information necessary for their reconstruction by the decoder; for this reason, they are the essential entry point for access to a video sequence. The compression rate of I pictures is relatively low, and is comparable to a JPEG coded picture of a similar resolution.
- **P (predicted) pictures** are coded from the preceding I or P picture, using the techniques of motion compensated prediction. P pictures can be used as the basis for next predicted pictures, but since motion compensation is not perfect, it is not possible to extend very much the number of P pictures between two I pictures. The compression rate of P pictures is significantly higher than for I pictures.
- **B (bidirectional or bidirectionally predicted) pictures** are coded by bidirectional interpolation between the I or P picture which precedes and follows them. As they are not used for coding

Fig. 2.1a: Arrangement of Frames in MPEG

- *Intra (I*) pictures are coded without reference to other pictures, in a very similar manner to JPEG, which means that they contain all the information necessary for their reconstruction by the decoder; for this reason, they are the essential entry point for access to a video sequence. The compression rate of *I* pictures is relatively low, and is comparable to a JPEG coded picture of a similar resolution.

- *Predicted (P)* pictures are coded from the preceding *I* or *P* picture, using the technique of motion compensated prediction. *P* pictures can be used as the basis for next predicted pictures, but due to inherent imperfection of the *motion compensation technique,* the number of *P* pictures between consecutive *I* pictures must of a necessity be limited. It follows logically that the compression rate of *P* pictures is significantly higher than that of *I* picture.
- *Bidirectional or bidirectionally predicted (B)* pictures are coded by bidirectional interpolation between I and P pictures preceding and following them. As they are not used for coding subsequent pictures, *B* pictures do not propagate coding errors. *B* pictures offer the highest compression rate.
- Depending on the complexity of the encoder used, it is possible to encode I only, I and P, or I, P and B pictures, with very different results with regards to compression rate and random access resolution, and also with regard to encoding time and perceived quality.

Two parameters; M and N, describe the succession of I, P and B frame as depicted in figure2.1b

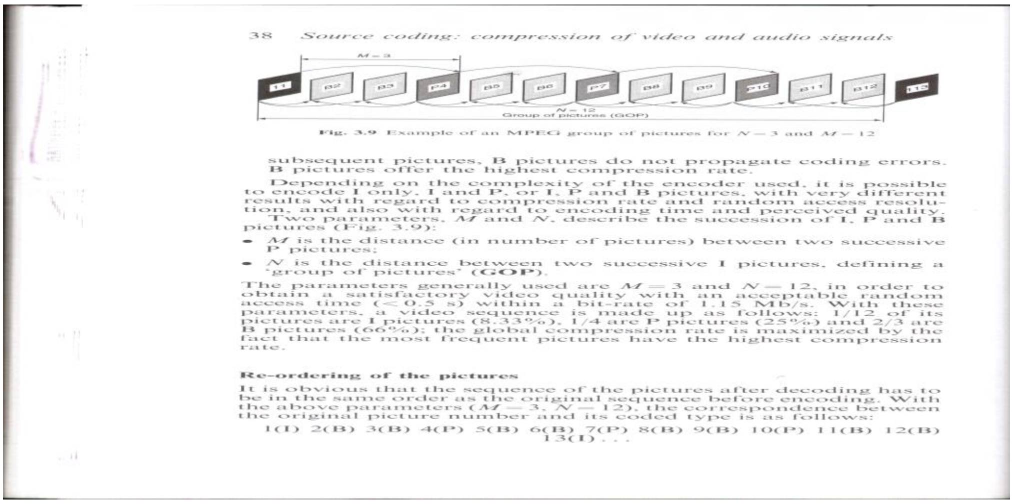

38 *Source coding: compression of video and audio signals*

Fig. 3.9 Example of an MPEG group of pictures for *N* = 3 and *M* = 12

subsequent pictures, B pictures do not propagate coding errors. B pictures offer the highest compression rate.

Depending on the complexity of the encoder used, it is possible to encode I only, I and P, or I, P and B pictures, with very different results with regard to compression rate and random access resolution, and also with regard to encoding time and perceived quality.

Two parameters, *M* and *N*, describe the succession of I, P and B pictures (Fig. 3.9):

- *M* is the distance (in number of pictures) between two successive P pictures;
- *N* is the distance between two successive I pictures, defining a 'group of pictures' (**GOP**).

The parameters generally used are *M* = 3 and *N* = 12, in order to obtain a satisfactory video quality with an acceptable random access time (< 0.5 s) within a bit-rate of 1.15 Mb/s. With these parameters, a video sequence is made up as follows: 1/12 of its pictures are I pictures (8.33%), 1/4 are P pictures (25%) and 2/3 are B pictures (66%); the global compression rate is maximized by the fact that the most frequent pictures have the highest compression rate.

Re-ordering of the pictures

It is obvious that the sequence of the pictures after decoding has to be in the same order as the original sequence before encoding. With the above parameters (*M* = 3, *N* = 12), the correspondence between the original picture number and its coded type is as follows:

1(I) 2(B) 3(B) 4(P) 5(B) 6(B) 7(P) 8(B) 9(B) 10(P) 11(B) 12(B) 13(I) . . .

Fig. 2.1b Example of an MPEG Group of Pictures for M=3 and N=12

M is the distance (in number of pictures) between two successive P pictures;

N is the distance between two successive I frames, defining a “group of pictures” (GOP).

The parameters generally used are M=3, and N=12, in order to obtain a satisfactory video quality with an acceptable random access time (less than 0.5s) within a bit-rate of 1.15 Mbps. With these parameters, a video sequence is made up as follows: 1/12 (i.e. 8.33%) of its pictures are I pictures, ¼ (25%) are P pictures, and 2/3 (66.67%) are B pictures. The resultant compression rate is maximised by the fact that the most frequent pictures have the highest compression rate.

Re-Ordering of the Picture / frame

It is obvious that the sequence of the picture after decoding has to be in the same order as the original sequence before encoding. With the parameters earlier stated (M=3, N=12), the difference between the original picture number and its coded type is as follows:

1(I) 2(B) 3 (B) 4 (P) 5 (B) 6 (B) 7 (P) 8 (B) 9 (B) 10 (P) 11 (B) 12(B) 13 (I)

However, in order to encode or decode bidirectional picture, both the encoder and the decoder will need the I or P preceding picture and the I or P subsequent picture. This requires re-ordering of the original picture sequence such that the decoder and the encoder have at their disposal the required I and (or) P picture before the B picture is processed. The re-ordering thus gives the following sequence:

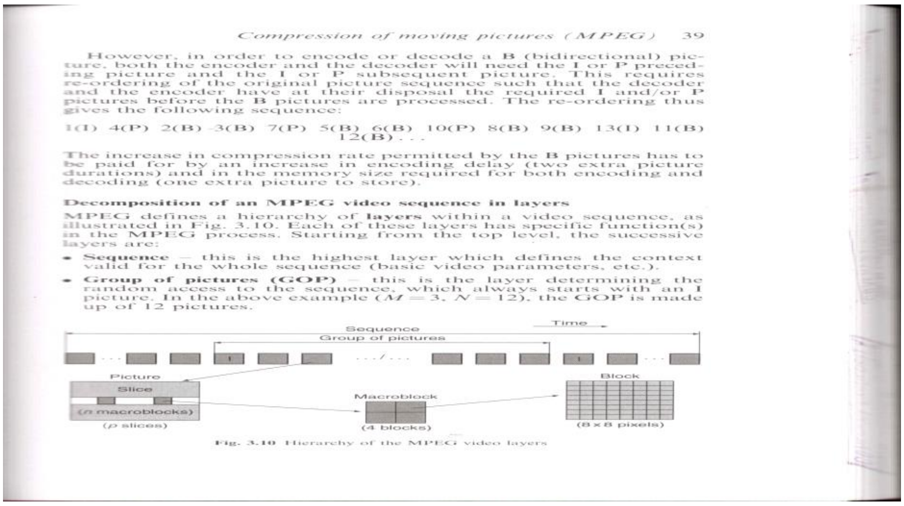
1(I), 4 (P), 2 (B) , 3(B), 7(P), 5 (B), 6 (B), 10 (P), 8(B), 9(B), 13 (I), 11 (B) 12(B) ..

The increase in compression rate introduced by the B picture has to be compensated for by an increase in encoding delay and in the memory size required for both encoding and decoding (one extra picture to store).

3.2.2 Decomposition of an MPEG Video Sequence in Layers

MPEG defines a hierarchy of layers within a video sequence, as shown in Figure 2.2a. Each of these layers has specific function(s) in the MPEG process. Starting from the top level, the successive layers are:

- *Sequence* – this is the highest layer defining the context valid for the whole sequence (basic video parameters, etc)
- *Group of Pictures (GOP)* - this layer determines the random access to the sequence, which always starts with an I picture. In the example in Figure 2.1b the GOP is made up of 12 pictures.


Compression of moving pictures (MPEG) 39

However, in order to encode or decode a B (bidirectional) picture, both the encoder and the decoder will need the I or P preceding picture and the I or P subsequent picture. This requires re-ordering of the original picture sequence such that the decoder and the encoder have at their disposal the required I and/or P pictures before the B pictures are processed. The re-ordering thus gives the following sequence:

1(I) 4(P) 2(B) 3(B) 7(P) 5(B) 6(B) 10(P) 8(B) 9(B) 13(I) 11(B) 12(B) . . .

The increase in compression rate permitted by the B pictures has to be paid for by an increase in encoding delay (two extra picture durations) and in the memory size required for both encoding and decoding (one extra picture to store).

Decomposition of an MPEG video sequence in layers

MPEG defines a hierarchy of **layers** within a video sequence, as illustrated in Fig. 3.10. Each of these layers has specific function(s) in the MPEG process. Starting from the top level, the successive layers are:

- **Sequence** – this is the highest layer which defines the context valid for the whole sequence (basic video parameters, etc.).
- **Group of pictures (GOP)** – this is the layer determining the random access to the sequence, which always starts with an I picture. In the above example ($M = 3$, $N = 12$), the GOP is made up of 12 pictures.

Fig. 3.10 Hierarchy of the MPEG video layers

Fig. 2.2a: Hierarchy of the MPEG VIDEO Layer

- *Picture* – this is also referred to as frame. It is the elementary display unit, which can be one of the three types (I, P, or B) discussed earlier.
- *Slice* – this is the layer for intra frame addressing, and (re)synchronisation, for instance for error recovery. It consists of a suite of contiguous macroblocks. Theoretically, the size of a slice can range from one macroblock to the whole picture, but in most cases it is made up of a complete row of macroblocks
- *Macroblocks* – this is the layers used for movement estimation / compensation. It consists of a size of 16 x 16 pixels and made up of four blocks of luminance and two blocks of chrominance. See figure 2.2b for more details.

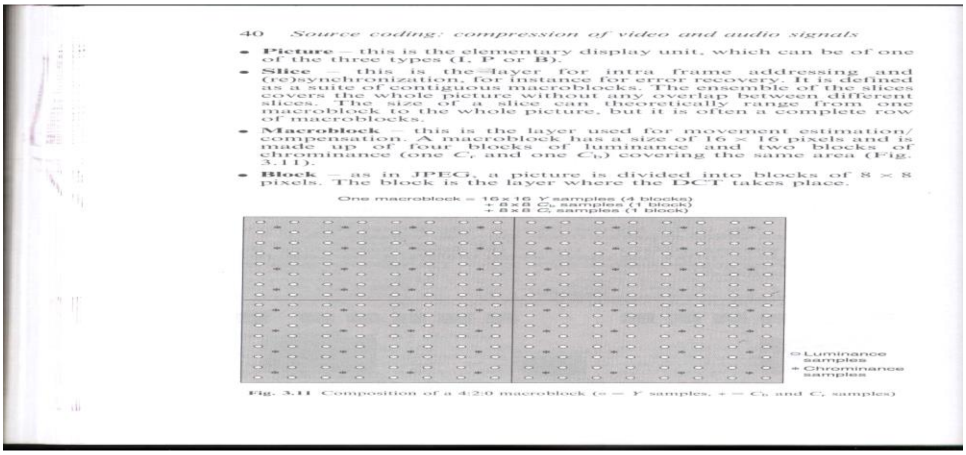
- *Block* – Just as in JPEG, a picture is made up of blocks of 8 x 8 pixels. It is the layer where the DCT takes place.

One macroblock = 16 x 16 Y samples (4 blocks)

+8 x 8 C_b samples (1 block)

+8 x 8 C_r samples (1 block)

0 =luminance sample, * Chrominance sample


40 *Source coding: compression of video and audio signals*

- **Picture** – this is the elementary display unit, which can be of one of the three types (I, P or B).
- **Slice** – this is the layer for intra frame addressing and (re)synchronization, for instance for error recovery. It is defined as a suite of contiguous macroblocks. The ensemble of the slices covers the whole picture without any overlap between different slices. The size of a slice can theoretically range from one macroblock to the whole picture, but it is often a complete row of macroblocks.
- **Macroblock** – this is the layer used for movement estimation/compensation. A macroblock has a size of 16 × 16 pixels and is made up of four blocks of luminance and two blocks of chrominance (one C_r and one C_b) covering the same area (Fig. 3.11).
- **Block** – as in JPEG, a picture is divided into blocks of 8 × 8 pixels. The block is the layer where the DCT takes place.

Fig. 3.11 Composition of a 4:2:0 macroblock (○ = Y samples, * = C_b and C_r samples)

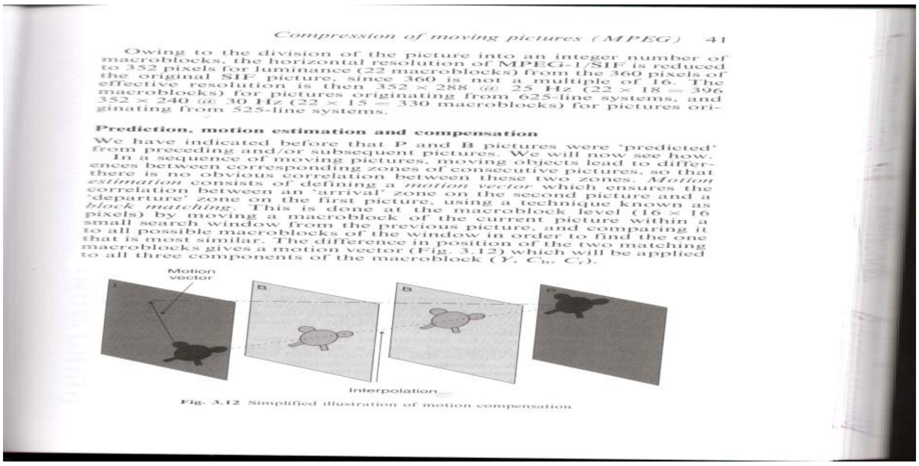
Fig. 2.2b: Composition of a 4:2:0 Macroblock (0=Y Sample, * = Cb and Cr Sample

Source H. Benoit,H.(1997)

3.2.3 Prediction, Motion Estimation and Compensation

In section 3.2.1, we discussed that P and B pictures were "predicted" from the preceding and / or subsequent pictures. We shall elaborate more on this to enhance your understanding.

When dealing with videos and animations, the moving objects lead to differences between corresponding zones of consecutive pictures, so that there is no clear correlation between these two zones. **Motion estimation** analyses the video frames and calculates where objects are moved to. Instead of transmitting all the data needed to represent the new frame, only the information (i.e the vector or new position) needed to move the object transmitted. This is done at the macroblock level (16 x 16 pixels) by allowing a macroblock of the current picture within a small search window from the previous picture, and comparing it to all possible macroblocks of the window in order to find the one that is most similar. The difference in position of the two matching macroblocks gives a motion vector which will be applied to all three component of the macroblock (Y, C_b, C_r). See figure 2.3 for details.



Compression of moving pictures (MPEG) 41

Owing to the division of the picture into an integer number of macroblocks, the horizontal resolution of MPEG-1/SIF is reduced to 352 pixels for luminance (22 macroblocks) from the 360 pixels of the original SIF picture, since 360 is not a multiple of 16. The effective resolution is then 352 × 288 @ 25 Hz (22 × 18 = 396 macroblocks) for pictures originating from 625-line systems, and 352 × 240 @ 30 Hz (22 × 15 = 330 macroblocks) for pictures originating from 525-line systems.

Prediction, motion estimation and compensation

We have indicated before that P and B pictures were 'predicted' from preceding and/or subsequent pictures. We will now see how.

In a sequence of moving pictures, moving objects lead to differences between corresponding zones of consecutive pictures, so that there is no obvious correlation between these two zones. *Motion estimation* consists of defining a *motion vector* which ensures the correlation between an 'arrival' zone on the second picture and a 'departure' zone on the first picture, using a technique known as *block matching*. This is done at the macroblock level (16 × 16 pixels) by moving a macroblock of the current picture within a small search window from the previous picture, and comparing it to all possible macroblocks of the window in order to find the one that is most similar. The difference in position of the two matching macroblocks gives a motion vector (Fig. 3.12) which will be applied to all three components of the macroblock (Y, C_B, C_R).

Fig. 3.12 Simplified illustration of motion compensation

Interpolation

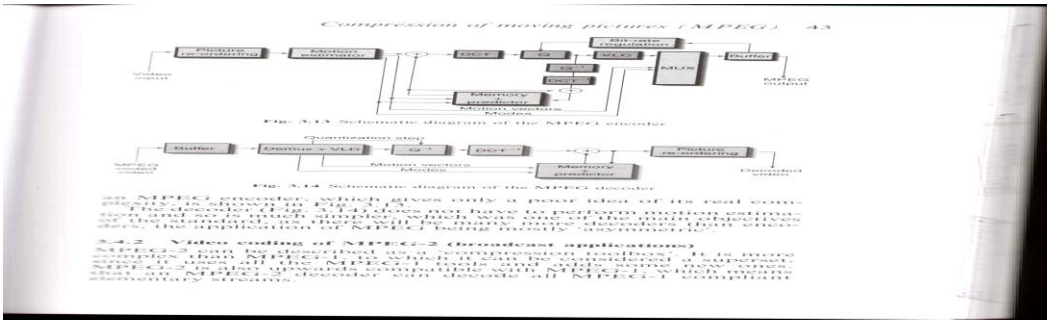
Fig. 2.3: A Simplified Illustration of Motion Compensation

Source Benoit,H. (1997)

Block matching is usually not adequate for making comparison in the temporary distance between a P picture and an I picture or two P picture as the motion vector can be of relatively high amplitude. Thus the difference (or prediction error) between the actual block to be encoded and the matching block has to be calculated, and encoded in a similar way to the block of the I pictures (DCT, quantisation, RLC/VLC). This process is called **motion compensation**.

In the case of B frames /pictures, motion vector are computed by temporal interpolation of the vector of the next P picture in three different ways (forward, backward and bidirectional); the smallest prediction error value is retained, and the error is encoded in the same way as for P pictures. Only the macroblocks which are different from the picture(s) used for prediction would need to be encoded. This process no doubt, helps to substantially reduce the amount of information required for coding B and P pictures. As the size of the moving objects is generally bigger than a macroblock, there is a strong correlation between the motion vectors of consecutive blocks, and a differential coding method (DPCM) is used to encode the vectors, thus reducing the number of bits required. In case, the "prediction" does not yield a desire result (for example, in the case of a moving camera where completely new zones appear in the picture), the corresponding parts of the picture are "Intra" coded in the same way as for I pictures. Figure 2.3a and Figure 2.3b depict the schematic diagrams for an MPEG encoder and decoder respectively. As will be observed, the decoder does not perform motion estimation and so is not as complex as an encoder. This is one of the main objectives of the standard.

Fig. 2.3a:



Schematic Diagram of the MPEG Encoder

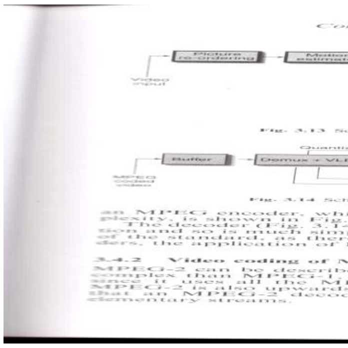


Fig. 2.3b: Schematic Diagram of the MPEG Decoder

4.0 Conclusion

Video compression typically operates on square-shaped groups of neighboring pixels, often called macroblocks. These pixel groups or blocks of pixels are compared from one frame to the next and the video compression codec sends only the differences within those blocks. This works extremely well if the video has no motion. A still frame of text, for example, can be repeated with very little transmitted data. In areas of video with more motion, more pixels change from one frame to the next. When more pixels change, the video compression scheme must send more data to keep up with the larger number of pixels that are changing.

5.0 Summary

In this unit, we learnt about video coding with motion estimation. Motion estimation involves comparing small segments of two consecutive frames for differences and, should a difference be detected, a search is carried out to determine to which neighbouring segment the original segment has moved. In order to minimise the time for each search, the search region is limited to just a few neighbouring segments. The MPEG technology employs this technique for data compression. Also in this unit, we covered motion compensation and prediction. The different types of frame and the MPEG video sequence layers were discussed.

6.0 Self-Assessment Exercise

1. Describe the different types of MPEG frames
2. Explain the term "Motion estimation"
3. With the aid of a labelled diagram, describe the MPEG the hierarchy of layers within MPEG video sequence

7.0 References/Further Reading

Benoit,H. (1997). Digital Television MPEG-1, MPEG-2 and Principles of the DVB systems.Oxford: Focal Press, Linacre House, Jordan Hill.

Gibson,J.D. (Ed.). (2001). Multimedia Communications: Directions and

Innovations.USA: Academic Press.

Unit 3 Image/Video Compression Standards Jpeg, Mpeg and H.263

1.0 Introduction

Image and video compression have become an integrated part of today's digital communications systems such as facsimile, videoconferencing, image archival systems, DVD, movie and video distribution, graphics and film industry, etc. As new application continues to emerge it is necessary to define standards for common data storage, compressions, retrieval, and transmission in these systems. This is to allow for perfect interoperability of data exchange.

The two main international bodies in the multimedia compression area are the International Organisation for Standardisation (ISO) and International Telecommunications Union – Telecommunications Sector (ITU-T) formerly known as Comité Consultatif International Téléphonique et Télégraphique (CCITT). In the following section we shall discuss the standards these two organisation support

2.0 Objectives

At the end of this unit, you should be able to:

- discuss an overview of different image / video standards
- explain important features of some common standards used in multimedia applications
- highlight the areas of applications of the standards

3.0 Main Content

3.1 Image/Video Compression Standard

The 'Joint Photographic Experts Group' or JPEG standard established by ISO (International Standards Organization) and IEC (International Electro-Technical Commission) is the standard for still image compression. MPEG (Moving Picture Expert Group) is the standard in ISO for digital compression system to handle pictures (video) and associated audio.

3.2 JPEG/ JPEG 2000

JPEG is the acronym for Joint Photographic Experts Group. It is the first international image compression standard for continuous-tone still images, including both gray scale and colour images. The goal of this standard is to support a variety of applications for compression of continuous-tone still images

(i) of different sizes,
(ii) in any colour space,
(iii) in order to achieve compression performance at or near the state of the art,
(iv) with user-adjustable compression ratios, and

(v) with very good to excellent reconstructed quality.

Another goal of this standard is that it should have manageable computational complexity for widespread practical implementation. JPEG defines the following four modes of operation

1. Sequential Lossless Mode: Compresses images in a single scan, and the decoded image is an exact replica of the original image.

2. Sequential DCT-Based Mode: Compresses images in a single scan using DCT-based lossy compression techniques. This gives a decoded image that is not an exact replica but an approximation of the original image.

3. Progressive DCT-Based Mode: This allows for the compression of images in multiple scans and also decompresses the image in multiple scans, with each successive scan producing a better quality image.

4. Hierarchical Mode: This allows for the compression of images at multiple resolutions for display on different devices.

The three DCT-based modes (2, 3, and 4) in JPEG provide lossy compression, because the precision limitation to digitally compute DCT (and its inverse) and the quantisation process introduce distortion in the reconstructed image. For sequential lossless mode of compression, predictive coding (DPCM) is used instead of the DCT-based transformation and also there is no quantisation involved.

The commonly used algorithm for image compression in the sequential DCT-based mode of the standard is called baseline JPEG. Great research efforts have been devoted into still image compression since the establishment of the JPEG standard in 1992. The success of JPEG necessitated further research efforts into an enhanced standard called JPEG-2000 for coding of still images. JPEG 2000 is a wavelet-based image compression standard. It was created by the Joint Photographic Experts Group committee in the year 2000 with the intention of superseding their original discrete cosine transform-based JPEG standard (created 1992). The standard include many modern features such as improved low bit-rate compression performance, lossless and lossy compression, continuous-tone and bi-level compression, compression of large images, single decompression architecture, transmission in noisy environments including robustness to bit-errors, progressive transmission by pixel accuracy and resolution, content-based description, and protective image security. The standardised filename extension is .jp2 for ISO/IEC 15444-1 conforming files and .jpx for the extended part-2 specifications, published as ISO/IEC 15444-2. The registered MIME types are defined in RFC 3745. For ISO/IEC 15444-1 it is image/jp2.

In addition, while there is a modest increase in compression performance of JPEG2000 compared to JPEG, another advantages offered by JPEG2000 is the significant flexibility of the codestream. The codestream obtained after compression of an image with JPEG2000 is scalable in nature, meaning that it can be decoded in a number of ways; for instance, by truncating the codestream at any point, one may obtain a representation of the image at a lower resolution. By ordering the codestream in various ways, applications can achieve significant performance increases. However, as a consequence of this flexibility, JPEG2000 requires encoders/decoders that are complex and computationally demanding. Another difference, in comparison with JPEG, is in terms of visual artifacts: JPEG 2000 produces ringing artifacts, manifested as blur and rings near edges in the image, while JPEG produces ringing artifacts and 'blocking' artifacts, due to its 8×8 blocks.

JPEG 2000 has been published as an ISO standard, ISO/IEC 15444. As of 2009, JPEG 2000 is not widely supported in web browsers, and hence is not generally used on the World Wide Web.

3.3 MPEG (Moving Picture Experts Group)

The Moving Picture Experts Group (MPEG) was formed by the ISO to set standards for audio and video compression and transmission. The MPEG standards consist of different Parts. Each part covers a certain aspect of the whole specification. The standards also specify Profiles and Levels. Profiles are intended to define a set of tools that are available, and Levels define the range of appropriate values for the properties associated with them. Some of the approved MPEG standards were revised by later amendments and /or new editions. MPEG has standardised the following compression formats and ancillary standards

3.3.1 MPEG-1

(officially known as ISO 11172) is the first generation of digital compression standards for video and two-channel stereo audio to achieve bit-rate of about 1.5 Mbps (Mega bits per seconds) for storage in CD-ROMs. MPEG-1 was standardized in 1994. This standard was based on CD-ROM video applications, and is a popular standard for video on the Internet, transmitted as .mpg files. In addition, level 3 of MPEG-1 is the most popular standard for digital compression of audio-known as MP3. MPEG-1 is the standard of compression for Video CD, the most popular video distribution format throughout Asia.

3.3.2 MPEG -2

This is the standard on which Digital Television set top boxes and DVD compression is based. It is based on MPEG-1, but designed for the compression and transmission of digital broadcast television. The most significant enhancement from MPEG-1 is its ability to efficiently compress interlaced video. MPEG-2 scales well to HDTV resolution and bit rates, obviating the need for an MPEG-3. MPEG-2 standard was considerably broader in scope and of wider appeal – supporting interlacing and high definition. MPEG-2 is considered important because it has been chosen as the compression scheme for over-the-air digital television ATSC, DVB and ISDB, digital satellite TV services like Dish Network, digital cable television signals, SVCD, and DVD. This is defined in a series of documents which are all subset of ISO Recommendation 13818. It is intended for the recording and transmission of studio-quality audio and video.

3.3.3 MPEG-4

MPEG-4 standard was defined to meet newer challenges of the object-based video coding suitable for multimedia applications. MPEG-4 is based on object-based compression, similar in nature to the Virtual Reality Modeling Language. Individual objects within a scene are tracked separately and compressed together to create an MPEG4 file. It allows developers to control objects independently in a scene, and therefore introduce interactivity. Initially, this standard was concerned with similar range of applications to those of H.263, each running over low bit rate channels ranging from 4.8 to 64 kps. Later it scope was expanded to embrace a wide range of interactive multimedia applications over the Internet and the various types of entertainment networks. The main application domain of the MPEG-4 standard is in relation to the audio and video associated with interactive multimedia applications over the Internet and the various types of entertainment network. The standard contains features to enable a user not only to passively access a video sequence (or complete video) – using, for example start/stop/pause command – but also to access and manipulate the individual elements that make up each scene within the sequence / video. If the accessed video is a computer-generated cartoon, for example, the user may be given the capability – by the creator of the video – to reposition, delete, or alter the movement of the individual characters within a scene. In addition, because of its high coding deficiency with scenes such as those associated video telephony, the standard is also used for this type of application running over low bit rate networks such as wireless and PSTNs. MPEG-4 uses further coding tools with additional complexity to achieve higher compression factors than MPEG-2. In addition to more efficient coding of video, MPEG-4 moves closer to computer graphics applications. In more complex profiles, the MPEG-4 decoder effectively becomes a rendering processor and the compressed bitstream describes three-dimensional shapes and surface texture. MPEG-4 also provides Intellectual Property Management and Protection (IPMP) which provides the facility to use proprietary technologies to manage and protect content like digital rights management. Several new higher-efficiency video standards (newer than MPEG-2 Video) are included (an alternative to MPEG-2 Video).

3.3.4 MPEG-7

There was a popular misconception that MPEG-7 was going to be another new video compression standard. The fact is that MPEG-7 does not define any new video compression standard. It deals with the file format and metadata description of the compressed video in order to define a standard for description of various types of multimedia coded with the standard codec. The main objective of MPEG-7 is to serve the need of audiovisual content-based retrieval (or audiovisual object retrieval) in applications such as digital libraries. Nevertheless, it is also applicable to any multimedia applications involving the generation and usage of multimedia data. MPEG-7 became an International Standard in September 2001 — with the formal name Multimedia Content Description Interface. MPEG-7 supports a variety of multimedia applications. Its data may include still pictures, graphics, 3D models, audio, speech, video, and composition information (how to combine these elements). These MPEG-7 data elements can be represented in textual format, or binary format, or both.

3.3.5 MPEG-21

The MPEG-21 standard established in 2001, from the Moving Picture Experts Group, aims at defining an open framework for multimedia applications. MPEG-21 is ratified in the standards ISO/IEC 21000 - Multimedia framework (MPEG-21)

Specifically, MPEG-21 defines a "Rights Expression Language" standard as means of sharing digital rights/permissions/restrictions for digital content from content creator to content consumer. As an XML-based standard, MPEG-21 is designed to communicate machine-readable license information and do so in a "ubiquitous, unambiguous and secure" manner.

MPEG-21 is based on two essential concepts: the definition of a fundamental unit of distribution and transaction, which is the Digital Item, and the concept of users interacting with them. Digital Items can be considered the kernel of the Multimedia Framework and the users can be considered as who interacts with them inside the Multimedia Framework. At its most basic level, MPEG-21 provides a framework in which one user interacts with another one, and the object of that interaction is a Digital Item. Due to that, we could say that the main objective of the MPEG-21 is to define the technology needed to support users to exchange, access, consume, trade or manipulate Digital Items in an efficient and transparent way.

3.3.6 Other MPEG Standards

Relatively more recently than other standards above, MPEG has started following international standards; each of the standards holds multiple MPEG technologies for a way of application. For example:

- MPEG-A (2007): Multimedia application format (MPEG-A). (ISO/IEC 23000) (e.g. Purpose for multimedia application formats, MPEG music player application format, MPEG photo player application format and others)
- MPEG-B (2006): MPEG systems technologies. (ISO/IEC 23001) (e.g. Binary MPEG format for XML, Fragment Request Units, Bitstream Syntax Description Language (BSDL) and others)
- MPEG-C (2006): MPEG video technologies. (ISO/IEC 23002) (e.g. Accuracy requirements for implementation of integer-output 8x8 inverse discrete cosine transform and others)
- MPEG-D (2007): MPEG audio technologies. (ISO/IEC 23003) (e.g. MPEG Surround and two parts under development: SAOC-Spatial Audio Object Coding and USAC-Unified Speech and Audio Coding)
- MPEG-E (2007): Multimedia Middleware. (ISO/IEC 23004) (a.k.a. M3W) (e.g. Architecture, Multimedia application programming interface (API), Component model and others)
- Supplemental media technologies (2008). (ISO/IEC 29116) Part 1: Media streaming application format protocols will be revised in MPEG-M Part 4 - MPEG extensible middleware (MXM) protocols.
- MPEG-V (under development at the time of writing this study material): Media context and control. (ISO/IEC CD 23005) (a.k.a. Information exchange with Virtual Worlds) (e.g. Avatar characteristics, Sensor information, Architecture and others)
- MPEG-M (under development at the time of writing this study material): MPEG extensible Middleware (MXM). (ISO/IEC FCD 23006) (e.g. MXM architecture and technologies, API, MPEG extensible middleware (MXM) protocols)
- MPEG-U (under development at the time of writing this study material): Rich media user interfaces. (ISO/IEC CD 23007) (e.g. Widgets)

3.4 H.263

H.263 is a video codec standard originally designed as a low-bit rate compression format for videoconferencing. It was developed by the ITU-T Video Coding Experts Group (VCEG) in a project ending in 1995/1996 as one member of the H.26x family of video coding standards in the domain of the ITU-T. H.263 was developed as an evolutionary improvement based on experience from H.261, the previous ITU-T standard for video compression, and the MPEG-1 and MPEG-2 standards.

Briefly, H.261 is an ITU standard designed for two-way communication over ISDN lines (video conferencing) and supports data rates which are multiples of 64Kbit/s. The algorithm is based on DCT and can be implemented in hardware or software and uses intraframe compression. H.261 supports CIF and QCIF resolutions.

The first version of H.263 was completed in 1995 and provided a suitable replacement for H.261 at all bitrates. It was further enhanced in projects known as H.263v2 (also known as H.263+ or H.263 1998) and H.263v3 (also known as H.263++ or H.263 2000).

The next enhanced codec developed by ITU-T VCEG (in partnership with MPEG) after H.263 is the H.264 standard, also known as AVC and MPEG-4 part 10. As H.264 provides a significant improvement in capability beyond H.263, the H.263 standard is now considered primarily a legacy design (although this is a recent development).

Most new videoconferencing products now include H.264 as well as H.263 and H.261 capabilities. The H.263 video compression standard has been defined by the ITU-T for use in a range of video applications over wireless and public switch telephone networks and in applications which include video telephony, videoconferencing, security surveillance, interactive game playing, the internet: much Flash Video content (as used on sites such as YouTube, Google Video, MySpace, etc.) used to be encoded in this format, though many sites now use VP6 or H.264 encoding. The original version of the RealVideo codec was based on H.263 up until the release of Real Video 8. H.263 is a required video codec in ETSI 3GPP technical specifications for IP Multimedia Subsystem (IMS), Multimedia Messaging Service (MMS) and Transparent end-to-end Packet-switched Streaming Service (PSS). In 3GPP specifications is H.263 video usually used in 3GP container format.

Self-Assessment Exercise

1. List five (5) electronic devices that support MPEG data format.
2. List two other formats apart from JPEG that you can use to store your image files.

Answer to Question 2
PNG – Portable Network Graphics
GIF - Graphics Interchange Format

4.0 Conclusion

The MPEG video standards are developed by experts in video compression working under the auspice of the International Organisation for Standardisation (ISO). The standards activity began in 1989 with the goal of developing a standard for a video compression algorithm suitable for use in CD-ROM based applications. The committee has since standardised MPEG-1, MPEG-2, MPEG-4, MPEG-7, and MPEG-21. JPEG is the acronym for Joint Photographic Experts Group. It is the first international image compression standard for continuous-tone still images, including both gray scale and colour images. The success of JPEG necessitated further research efforts into an improved standard called JPEG-2000 for coding of still images. JPEG 2000 is a wavelet-based image compression standard. H.263 is a video codec standard originally designed as a low-bit rate compressed format for videoconferencing. It was developed by the ITU-T Video Coding Experts Group (VCEG). It is one of the standards in the H.26x family of video coding in the domain of the ITU-T.

5.0 Summary

We learnt about some data compression standards for multimedia elements – images, video, speech, audio, text etc. The standards discussed in this unit will allow for interoperability of multimedia data across various systems or application domain. These standards are; JPEG/JPEG2000, MPEG, and the H.263 standard. While discussing H.263 standard we compared it with H.261, its predecessor H.264, and H.26x family in order to have a better understanding of its features.

6.0 Self –Assessment Exercise

1. Describe the four operations modes of JPEG
2. What are the main extensions in MPEG 4 compared with MPEG 2
3. Briefly discuss the features of H.263 standard

7.0 References/Further Reading

Acharya, T., -Sing Tsai,P. (2005). *JPEG2000 Standard for Image Compression*: Concept, Algorithm and VLSI.

Benoit,H.(1997). *Digital Television, MPEG-1, MPEG-2 and Principles of the DVB System*.Oxford: Focal Press, An Imprint of Elsevier, Jordan Hill.

Forouzan,B.A., Fegan,S.C. (2003). *Data Communication and Networking*. Singapore: McGraw Hill Higher Education.

Furht,B.,*et al.* (1995). *Video and Image Processing in Multimedia Systems*. Kluwer Academic Publisher.

Mitra,S.,& Tinkuacharya, (2003). *Data Mining Multimedia, Soft Computing, and Bioinformatics*.Canada: John Wiley & Sons, Inc, Hoboken. .

Strannedby,D., Walker,W.(2004). *Digital Signal Processing and Applications*.Oxford: Newness, An imprint of Elsevier, Jordan Hill.

http://www.mpeg.org/

http://www.jpeg.org/